Bayesian Inference with Sets of Conjugate Priors: Parameter Set Shapes and Model Behaviour

Gero Walter

Department of Statistics
Ludwig-Maximilians-Universität München (LMU)

March 6th, 2013

Outline

1. Common-cause failure modelling
   (joint work with Matthias Troffaes and Dana Kelly)
2. Generalised Bayesian inference with sets of priors
   (joint work with Thomas Augustin)
3. Prior-data conflict and Strong prior-data agreement
   (joint work with Thomas Augustin and Frank Coolen)

Common-Cause Failures

Source: Wikimedia Commons, `http://commons.wikimedia.org/wiki/File:Fukushima_I_by_Digital_Globe.jpg`

Common-Cause Failures

Source: http://www.diakont.com/solutions/nuclear-energy/plant-systems/diesel-generator-control-systems/

Source: Wikimedia Commons, http://commons.wikimedia.org/wiki/File:Fukushima_I_by_Digital_Globe.jpg

Common-Cause Failures

- All 12 generators (for 6 reactors) at Fukushima Daiichi were not available due to flooding of machine rooms (Tsunami caused by Tōhoku earthquake)

Common-Cause Failures

- All 12 generators (for 6 reactors) at Fukushima Daiichi were not available due to flooding of machine rooms (Tsunami caused by Tōhoku earthquake)

common-cause failure

simultaneous failure of several redundant components due to a common or shared root cause (Høyland & Rausand, 1994)

Common-Cause Failures

- All 12 generators (for 6 reactors) at Fukushima Daiichi were not available due to flooding of machine rooms (Tsunami caused by Tōhoku earthquake)

common-cause failure

simultaneous failure of several redundant components due to a common or shared root cause (Høyland & Rausand, 1994)

- Reliability of redundant systems

Common-Cause Failures

- All 12 generators (for 6 reactors) at Fukushima Daiichi were not available due to flooding of machine rooms (Tsunami caused by Tōhoku earthquake)

common-cause failure

simultaneous failure of several redundant components due to a common or shared root cause (Høyland & Rausand, 1994)

- Reliability of redundant systems
- Usually 2 – 4 emergency diesel generators per reactor

Common-Cause Failures

- All 12 generators (for 6 reactors) at Fukushima Daiichi were not available due to flooding of machine rooms (Tsunami caused by Tōhoku earthquake)

common-cause failure

simultaneous failure of several redundant components due to a common or shared root cause (Høyland & Rausand, 1994)

- Reliability of redundant systems
- Usually 2 – 4 emergency diesel generators per reactor
- Sufficient cooling of core if one generator works

Common-Cause Failures

- All 12 generators (for 6 reactors) at Fukushima Daiichi were not available due to flooding of machine rooms (Tsunami caused by Tōhoku earthquake)

common-cause failure

simultaneous failure of several redundant components due to a common or shared root cause (Høyland & Rausand, 1994)

- Reliability of redundant systems
- Usually 2 – 4 emergency diesel generators per reactor
- Sufficient cooling of core if one generator works
- Redundant components may not fail independently: common-cause failure

Common-Cause Failures

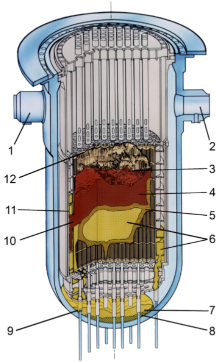
- All 12 generators (for 6 reactors) at Fukushima Daiichi were not available due to flooding of machine rooms (Tsunami caused by Tōhoku earthquake)

common-cause failure

simultaneous failure of several redundant components due to a common or shared root cause (Høyland & Rausand, 1994)

- Reliability of redundant systems
- Usually 2 – 4 emergency diesel generators per reactor
- Sufficient cooling of core if one generator works
- Redundant components may not fail independently: common-cause failure
- Must include common-cause failures in overall system reliability analysis

Common-Cause Failure Modelling



Above: CDC, http://phil.cdc.gov/phil/ ID 1194

Right: Wikimedia Commons,
http://commons.wikimedia.org/wiki/File:Graphic_TMI-2_Core_End-State_Configuration.png

Alpha-Factor Model: Definition

Alpha-Factor Model

Multinomial distribution $\mathrm{M}(\boldsymbol{n} \mid \boldsymbol{\alpha})$ for common-cause failures in a k-component system

$$p(\boldsymbol{n} \mid \boldsymbol{\alpha}) = \prod_{j=1}^{k} \alpha_j^{n_j}$$

where

- alpha-factor α_j := probability of j of the k components failing due to a common cause given that failure occurs
- failure count n_j := corresponding number of failures observed
- $\boldsymbol{n}$ denotes $(n_1, \ldots, n_k)$ and $\boldsymbol{\alpha}$ denotes $(\alpha_1, \ldots, \alpha_k)$

(the model actually serves to estimate failure *rates*, but the above is all what matters in this talk)

Alpha-Factor Model: Parameter Estimation

The Good News

attractive feature of this model:
α can be estimated directly from data, e.g. MLE:

$$\alpha_j = \frac{n_j}{n}, \quad \text{where } \sum_{j=1}^{n} n_j = n$$

Alpha-Factor Model: Parameter Estimation

The Good News

attractive feature of this model:
α can be estimated directly from data, e.g. MLE:

$$\alpha_j = \frac{n_j}{n}, \quad \text{where } \sum_{j=1}^{n} n_j = n$$

The Bad News

- typically, for $j \geq 2$, the n_j are very low
  with zero being quite common for larger j
- zero counts = flat likelihoods
  standard techniques such as MLE can struggle
  to produce sensible inferences for this problem

Alpha-Factor Model: Parameter Estimation

The Good News

attractive feature of this model:
α can be estimated directly from data, e.g. MLE:

$$\alpha_j = \frac{n_j}{n}, \quad \text{where } \sum_{j=1}^{n} n_j = n$$

The Bad News

- typically, for $j \geq 2$, the n_j are very low
  with zero being quite common for larger j
- zero counts = flat likelihoods
  standard techniques such as MLE can struggle
  to produce sensible inferences for this problem

→ **need to rely on epistemic information**

Bayesian Analysis: Dirichlet Prior

α considered as uncertain parameter on which we put...

Dirichlet Distribution (→ Dirichlet-Multinomial Model)

$$p(\alpha \mid n^{(0)}, \mathbf{y}^{(0)}) \propto \prod_{j=1}^{k} \alpha_j^{n^{(0)} y_j^{(0)} - 1}$$

where $(n^{(0)}, \mathbf{y}^{(0)})$ are *hyperparameters*

$n^{(0)} > 0$

$$\mathbf{y}^{(0)} \in \Delta = \left\{ (y_1^{(0)}, \ldots, y_k^{(0)}) : y_1^{(0)} \geq 0, \ldots, y_k^{(0)} \geq 0, \sum_{j=1}^{k} y_j^{(0)} = 1 \right\}$$

Bayesian Analysis: Dirichlet Prior

α considered as uncertain parameter on which we put. . .

Dirichlet Distribution (→ Dirichlet-Multinomial Model)

$$p(\alpha \mid n^{(0)}, \mathbf{y}^{(0)}) \propto \prod_{j=1}^{k} \alpha_j^{n^{(0)} y_j^{(0)} - 1}$$

where $(n^{(0)}, \mathbf{y}^{(0)})$ are *hyperparameters*

$n^{(0)} > 0$

$$\mathbf{y}^{(0)} \in \Delta = \left\{ (y_1^{(0)}, \ldots, y_k^{(0)}) : y_1^{(0)} \geq 0, \ldots, y_k^{(0)} \geq 0, \sum_{j=1}^{k} y_j^{(0)} = 1 \right\}$$

Interpretation

- $\mathbf{y}^{(0)}$ = prior expectation of α, i.e., a prior guess for $\frac{n_j}{n}$, $j = 1, \ldots, n$
- $n^{(0)}$ = determines spread and learning speed (see next slide)

Bayesian Analysis: Dirichlet Posterior

- posterior density for $\boldsymbol{\alpha}$ is again Dirichlet ➡ conjugacy
  update parameters: $n^{(0)} \to n^{(n)}$, $\mathbf{y}^{(0)} \to \mathbf{y}^{(n)}$

$$p(\boldsymbol{\alpha} \mid n^{(0)}, \mathbf{y}^{(0)}, \mathbf{n}) = p(\boldsymbol{\alpha} \mid n^{(n)}, \mathbf{y}^{(n)}) \propto \prod_{j=1}^{k} \alpha_j^{n^{(n)} y_j^{(n)} - 1}$$

Bayesian Analysis: Dirichlet Posterior

- posterior density for $\boldsymbol{\alpha}$ is again Dirichlet ➡ conjugacy
  update parameters: $n^{(0)} \to n^{(n)}$, $\boldsymbol{y}^{(0)} \to \boldsymbol{y}^{(n)}$

$$p(\boldsymbol{\alpha} \mid n^{(0)}, \boldsymbol{y}^{(0)}, \boldsymbol{n}) = p(\boldsymbol{\alpha} \mid n^{(n)}, \boldsymbol{y}^{(n)}) \propto \prod_{j=1}^{k} \alpha_j^{n^{(n)} y_j^{(n)} - 1}$$

- posterior expectation of α_j:

$$\mathrm{E}[\alpha_j \mid n^{(0)}, \boldsymbol{y}^{(0)}, \boldsymbol{n}] = \mathrm{E}[\alpha_j \mid n^{(n)}, \boldsymbol{y}^{(n)}] = \int_{\Delta} \alpha_j p(\boldsymbol{\alpha} \mid n^{(0)}, \boldsymbol{y}^{(0)}, \boldsymbol{n}) \, \mathrm{d}\boldsymbol{\alpha}$$

$$= y_j^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y_j^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{n_j}{n}$$

Bayesian Analysis: Dirichlet Posterior

- posterior density for $\boldsymbol{\alpha}$ is again Dirichlet ➡ conjugacy
  update parameters: $n^{(0)} \to n^{(n)}$, $\mathbf{y}^{(0)} \to \mathbf{y}^{(n)}$

$$p(\boldsymbol{\alpha} \mid n^{(0)}, \mathbf{y}^{(0)}, \mathbf{n}) = p(\boldsymbol{\alpha} \mid n^{(n)}, \mathbf{y}^{(n)}) \propto \prod_{j=1}^{k} \alpha_j^{n^{(n)} y_j^{(n)} - 1}$$

- posterior expectation of α_j:

$$\mathrm{E}[\alpha_j \mid n^{(0)}, \mathbf{y}^{(0)}, \mathbf{n}] = \mathrm{E}[\alpha_j \mid n^{(n)}, \mathbf{y}^{(n)}] = \int_\Delta \alpha_j p(\boldsymbol{\alpha} \mid n^{(0)}, \mathbf{y}^{(0)}, \mathbf{n}) \,\mathrm{d}\boldsymbol{\alpha}$$

$$= y_j^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y_j^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{n_j}{n}$$

we will focus on $\mathrm{E}[\alpha_j \mid n^{(n)}, \mathbf{y}^{(n)}]$
(in a decision context, this expectation would typically end up in expressions for expected utility)

Example: Prior and Data

(taken from Kelly & Atwood, 2011)

Example

Consider a system with four redundant components ($k = 4$).
The analyst specifies the following prior expectation $\mu_{\text{spec},j}$ for each α_j:

$$\mu_{\text{spec},1} = 0.950 \quad \mu_{\text{spec},2} = 0.030 \quad \mu_{\text{spec},3} = 0.015 \quad \mu_{\text{spec},4} = 0.005$$

We have 36 observations, in which 35 showed one component failing, and 1 showed two components failing:

$$n_1 = 35 \qquad n_2 = 1 \qquad n_3 = 0 \qquad n_4 = 0$$

Example: Non-Informative Priors

large variation in posterior under different non-informative priors

- with constrained maximum entropy prior (Atwood, 1996; Kelly & Atwood, 2011):

$$E[\alpha_1 \mid n^{(n)}, y^{(n)}] = 0.967 \qquad E[\alpha_2 \mid n^{(n)}, y^{(n)}] = 0.028$$
$$E[\alpha_3 \mid n^{(n)}, y^{(n)}] = 0.003 \qquad E[\alpha_4 \mid n^{(n)}, y^{(n)}] = 0.001$$

Example: Non-Informative Priors

large variation in posterior under different non-informative priors

- with constrained maximum entropy prior (Atwood, 1996; Kelly & Atwood, 2011):

$$E[\alpha_1 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.967 \qquad E[\alpha_2 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.028$$
$$E[\alpha_3 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.003 \qquad E[\alpha_4 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.001$$

- with uniform prior $y_j^{(0)} = 0.25$ and $n^{(0)} = 4$:

$$E[\alpha_1 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.9 \qquad E[\alpha_2 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.05$$
$$E[\alpha_3 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.025 \qquad E[\alpha_4 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.025$$

Example: Non-Informative Priors

large variation in posterior under different non-informative priors

- with constrained maximum entropy prior (Atwood, 1996; Kelly & Atwood, 2011):

$$E[\alpha_1 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.967 \qquad E[\alpha_2 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.028$$
$$E[\alpha_3 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.003 \qquad E[\alpha_4 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.001$$

- with uniform prior $y_j^{(0)} = 0.25$ and $n^{(0)} = 4$:

$$E[\alpha_1 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.9 \qquad E[\alpha_2 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.05$$
$$E[\alpha_3 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.025 \qquad E[\alpha_4 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.025$$

- with Jeffrey's prior $y_j^{(0)} = 0.25$ and $n^{(0)} = 2$:

$$E[\alpha_1 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.9342 \qquad E[\alpha_2 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.0395$$
$$E[\alpha_3 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.0132 \qquad E[\alpha_4 \mid n^{(n)}, \mathbf{y}^{(n)}] = 0.0132$$

Imprecise Dirichlet Model: Definition

Troffaes, Walter & Kelly (2012): model vague prior info more cautiously

Imprecise Dirichlet Model (IDM) for Common-Cause Failure

use a set of hyperparameters (Walley 1991, 1996):

$$= \left\{ (n^{(0)}, \mathbf{y}^{(0)}) \colon n^{(0)} \in [\underline{n}^{(0)}, \overline{n}^{(0)}], \mathbf{y}^{(0)} \in \Delta, y_j^{(0)} \in [\underline{y}_j^{(0)}, \overline{y}_j^{(0)}] \right\}$$

Imprecise Dirichlet Model: Definition

Troffaes, Walter & Kelly (2012): model vague prior info more cautiously

Imprecise Dirichlet Model (IDM) for Common-Cause Failure

use a set of hyperparameters (Walley 1991, 1996):

$$= \left\{ (n^{(0)}, \mathbf{y}^{(0)}) \colon n^{(0)} \in [\underline{n}^{(0)}, \overline{n}^{(0)}],\ \mathbf{y}^{(0)} \in \Delta,\ y_j^{(0)} \in [\underline{y}_j^{(0)}, \overline{y}_j^{(0)}] \right\}$$

Interpretation

- we are doing a sensitivity analysis (á la robust Bayes) over $(n^{(0)}, \mathbf{y}^{(0)}) \in$
- we take a set of priors based on as model for prior information (details later)

Imprecise Dirichlet Model: Definition

Troffaes, Walter & Kelly (2012): model vague prior info more cautiously

Imprecise Dirichlet Model (IDM) for Common-Cause Failure

use a set of hyperparameters (Walley 1991, 1996):

$$= \left\{ (n^{(0)}, \mathbf{y}^{(0)}) \colon n^{(0)} \in [\underline{n}^{(0)}, \overline{n}^{(0)}],\ \mathbf{y}^{(0)} \in \Delta,\ y_j^{(0)} \in [\underline{y}_j^{(0)}, \overline{y}_j^{(0)}] \right\}$$

Interpretation

- we are doing a sensitivity analysis (á la robust Bayes) over $(n^{(0)}, \mathbf{y}^{(0)}) \in$
- we take a set of priors based on as model for prior information (details later)

Analyst has to specify
bounds $[\underline{n}^{(0)}, \overline{n}^{(0)}]$ and bounds $[\underline{y}_j^{(0)}, \overline{y}_j^{(0)}]$ for each $j \in \{1, \dots, k\}$

Imprecise Dirichlet Model: Elicitation

- $[\underline{y}_j^{(0)}, \overline{y}_j^{(0)}]$: Cautious interpretation of prior specifications $\mu_{\text{spec},j}$:

$$[\underline{y}_1^{(0)}, \overline{y}_1^{(0)}] = [0.950, 1] \qquad [\underline{y}_2^{(0)}, \overline{y}_2^{(0)}] = [0, 0.030]$$

$$[\underline{y}_3^{(0)}, \overline{y}_3^{(0)}] = [0, 0.015] \qquad [\underline{y}_4^{(0)}, \overline{y}_4^{(0)}] = [0, 0.005]$$

Imprecise Dirichlet Model: Elicitation

- $[\underline{y}_j^{(0)}, \overline{y}_j^{(0)}]$: Cautious interpretation of prior specifications $\mu_{\text{spec},j}$:

$$[\underline{y}_1^{(0)}, \overline{y}_1^{(0)}] = [0.950, 1] \qquad [\underline{y}_2^{(0)}, \overline{y}_2^{(0)}] = [0, 0.030]$$

$$[\underline{y}_3^{(0)}, \overline{y}_3^{(0)}] = [0, 0.015] \qquad [\underline{y}_4^{(0)}, \overline{y}_4^{(0)}] = [0, 0.005]$$

- $[\underline{n}^{(0)}, \overline{n}^{(0)}]$: Good (1965):

  reason about posterior expectations for hypothetical data

Imprecise Dirichlet Model: Elicitation

- $[\underline{y}_j^{(0)}, \overline{y}_j^{(0)}]$: Cautious interpretation of prior specifications $\mu_{\text{spec},j}$:

$$[\underline{y}_1^{(0)}, \overline{y}_1^{(0)}] = [0.950, 1] \qquad [\underline{y}_2^{(0)}, \overline{y}_2^{(0)}] = [0, 0.030]$$

$$[\underline{y}_3^{(0)}, \overline{y}_3^{(0)}] = [0, 0.015] \qquad [\underline{y}_4^{(0)}, \overline{y}_4^{(0)}] = [0, 0.005]$$

- $[\underline{n}^{(0)}, \overline{n}^{(0)}]$: Good (1965):

  reason about posterior expectations for hypothetical data

$\overline{n}^{(0)}$ = number of one-component failures required
to reduce the upper probabilities of multi-component failure by half

$\underline{n}^{(0)}$ = number of multi-component failures required
to reduce the lower probability of one-component failure by half

Imprecise Dirichlet Model: Elicitation

$\overline{n}^{(0)}$ = number of one-component failures required
to reduce the upper probabilities of multi-component failure by half

$\underline{n}^{(0)}$ = number of multi-component failures required
to reduce the lower probability of one-component failure by half

Imprecise Dirichlet Model: Elicitation

$\overline{n}^{(0)}$ = number of one-component failures required
to reduce the upper probabilities of multi-component failure by half

$\underline{n}^{(0)}$ = number of multi-component failures required
to reduce the lower probability of one-component failure by half

Reasonable values in example:

- $\underline{n}^{(0)} = 1$: immediate multi-component failure
  ➡ keen to reduce lower probability for one-component failure
- $\overline{n}^{(0)} = 10$: after observing 10 one-component failures
  ➡ halve upper probabilities of multi-component failures

Imprecise Dirichlet Model: Elicitation

$\overline{n}^{(0)}$ = number of one-component failures required
to reduce the upper probabilities of multi-component failure by half

$\underline{n}^{(0)}$ = number of multi-component failures required
to reduce the lower probability of one-component failure by half

Reasonable values in example:

- $\underline{n}^{(0)} = 1$: immediate multi-component failure
  ➡ keen to reduce lower probability for one-component failure
- $\overline{n}^{(0)} = 10$: after observing 10 one-component failures
  ➡ halve upper probabilities of multi-component failures

Difference between $\underline{n}^{(0)}$ and $\overline{n}^{(0)}$ reflects a level of caution:
The rate at which we reduce upper probabilities
is less than the rate at which we reduce lower probabilities

Imprecise Dirichlet Model: Inference

prior bounds + likelihood → posterior bounds

with $y_j^{(0)} = \mu_{\text{spec},j}$: | with bounds as earlier:

j	$\underline{\mathrm{E}}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\overline{\mathrm{E}}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\underline{\mathrm{E}}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\overline{\mathrm{E}}[\alpha_j \mid \quad , \boldsymbol{n}]$
1	0.967	0.972	0.967	0.978
2	0.0278	0.0283	0.0270	0.0283
3	0.00041	0.00326	0	0.00326
4	0.00014	0.00109	0	0.00109

Imprecise Dirichlet Model: Inference

prior bounds + likelihood → posterior bounds

with $y_j^{(0)} = \mu_{\text{spec},j}$: | with bounds as earlier:

j	$\underline{\mathrm{E}}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\overline{\mathrm{E}}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\underline{\mathrm{E}}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\overline{\mathrm{E}}[\alpha_j \mid \quad , \boldsymbol{n}]$
1	0.967	0.972	0.967	0.978
2	0.0278	0.0283	0.0270	0.0283
3	0.00041	0.00326	0	0.00326
4	0.00014	0.00109	0	0.00109

- Bounds, rather than precise values, are desirable due to inferences being strongly sensitive to the prior particularly when faced with zero counts

Imprecise Dirichlet Model: Inference

prior bounds + likelihood → posterior bounds

	with $y_j^{(0)} = \mu_{\text{spec},j}$:		with bounds as earlier:	
j	$\underline{E}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\overline{E}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\underline{E}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\overline{E}[\alpha_j \mid \quad , \boldsymbol{n}]$
1	0.967	0.972	0.967	0.978
2	0.0278	0.0283	0.0270	0.0283
3	0.00041	0.00326	0	0.00326
4	0.00014	0.00109	0	0.00109

- Bounds, rather than precise values, are desirable due to inferences being strongly sensitive to the prior particularly when faced with zero counts
- Simple ways to elicit the parameters of the model by reasoning on hypothetical data

Imprecise Dirichlet Model: Inference

prior bounds + likelihood → posterior bounds

with $y_j^{(0)} = \mu_{\text{spec},j}$: || with bounds as earlier:

j	$\underline{E}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\overline{E}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\underline{E}[\alpha_j \mid \quad , \boldsymbol{n}]$	$\overline{E}[\alpha_j \mid \quad , \boldsymbol{n}]$
1	0.967	0.972	0.967	0.978
2	0.0278	0.0283	0.0270	0.0283
3	0.00041	0.00326	0	0.00326
4	0.00014	0.00109	0	0.00109

- Bounds, rather than precise values, are desirable due to inferences being strongly sensitive to the prior particularly when faced with zero counts
- Simple ways to elicit the parameters of the model by reasoning on hypothetical data
- Is it possible to generalise this method to other problems?

Canonical Conjugate Priors

The multinomial is an example for a canonical exponential family

$(x_1, \ldots, x_n) = \boldsymbol{x} \overset{iid}{\sim}$ canonical exponential family

$$p(\boldsymbol{x} \mid \theta) \propto \exp\Big\{\langle \psi, \tau(\boldsymbol{x}) \rangle - nb(\psi)\Big\} \qquad \Big[\psi \text{ transformation of } \theta\Big]$$

(includes Binomial, Multinomial, Normal, Poisson, Exponential, ...)

Canonical Conjugate Priors

The multinomial is an example for a canonical exponential family

$(x_1, \ldots, x_n) = \boldsymbol{x} \overset{iid}{\sim}$ canonical exponential family

$$p(\boldsymbol{x} \mid \theta) \propto \exp\left\{\langle \psi, \tau(\boldsymbol{x}) \rangle - n b(\psi)\right\} \qquad \left[\psi \text{ transformation of } \theta\right]$$

(includes Binomial, Multinomial, Normal, Poisson, Exponential, . . .)

- conjugate prior: $p(\psi \mid n^{(0)}, y^{(0)}) \propto \exp\left\{n^{(0)}\left[\langle \psi, y^{(0)} \rangle - b(\psi)\right]\right\}$

Canonical Conjugate Priors

The multinomial is an example for a canonical exponential family

$(x_1, \ldots, x_n) = \boldsymbol{x} \overset{iid}{\sim}$ canonical exponential family

$$p(\boldsymbol{x} \mid \theta) \propto \exp\Big\{\langle \psi, \tau(\boldsymbol{x}) \rangle - n b(\psi)\Big\} \qquad \Big[\psi \text{ transformation of } \theta\Big]$$

(includes Binomial, Multinomial, Normal, Poisson, Exponential, . . .)

- conjugate prior: $p(\psi \mid n^{(0)}, y^{(0)}) \propto \exp\Big\{n^{(0)}\Big[\langle \psi, y^{(0)} \rangle - b(\psi)\Big]\Big\}$
- (conjugate) posterior: $p(\psi \mid n^{(0)}, y^{(0)}, \boldsymbol{x}) \propto \exp\Big\{n^{(n)}\Big[\langle \psi, y^{(n)} \rangle - b(\psi)\Big]\Big\}$

where $y^{(n)} = \dfrac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \dfrac{n}{n^{(0)} + n} \cdot \dfrac{\tau(\boldsymbol{x})}{n}$ and $n^{(n)} = n^{(0)} + n$

Canonical Conjugate Priors

- (conjugate) posterior: $p(\psi \mid n^{(n)}, y^{(n)}) \propto \exp\left\{n^{(n)}\left[\langle \psi, y^{(n)} \rangle - b(\psi)\right]\right\}$

where $y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{\tau(\boldsymbol{x})}{n}$ and $n^{(n)} = n^{(0)} + n$

Canonical Conjugate Priors

- (conjugate) posterior: $p(\psi \mid n^{(n)}, y^{(n)}) \propto \exp\left\{n^{(n)}\left[\langle\psi, y^{(n)}\rangle - b(\psi)\right]\right\}$

where $y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{\tau(\boldsymbol{x})}{n}$ and $n^{(n)} = n^{(0)} + n$

Interpretation

- $n^{(0)}$ = determines spread and learning speed
- $y^{(0)}$ = prior expectation of $\tau(\boldsymbol{x})/n$

Canonical Conjugate Priors

- (conjugate) posterior: $p(\psi \mid n^{(n)}, y^{(n)}) \propto \exp\left\{ n^{(n)} \left[\langle \psi, y^{(n)} \rangle - b(\psi) \right] \right\}$

where $y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{\tau(\boldsymbol{x})}{n}$ and $n^{(n)} = n^{(0)} + n$

Interpretation

- $n^{(0)}$ = determines spread and learning speed
- $y^{(0)}$ = prior expectation of $\tau(\boldsymbol{x})/n$

Example: Scaled Normal Data

Data :	$\boldsymbol{x} \mid \mu$	$\sim$	$\mathrm{N}(\mu, 1)$
conjugate prior:	$\mu \mid n^{(0)}, y^{(0)}$	$\sim$	$\mathrm{N}(y^{(0)}, 1/n^{(0)})$
posterior:	$\mu \mid n^{(n)}, y^{(n)}$	$\sim$	$\mathrm{N}(y^{(n)}, 1/n^{(n)})$ $\quad (\frac{\tau(\boldsymbol{x})}{n} = \bar{\boldsymbol{x}})$

Bayesian Inference with Sets of Priors

Standard Bayesian inference procedure

$$\text{prior} + \text{likelihood} = \text{posterior}$$

using Bayes' Rule
All inferences are based on the posterior
(e.g., point estimate = $E[\psi \mid n^{(n)}, y^{(n)}]$)

Bayesian Inference with Sets of Priors

Standard Bayesian inference procedure

prior + likelihood = posterior

using Bayes' Rule
All inferences are based on the posterior
(e.g., point estimate = $E[\psi \mid n^{(n)}, y^{(n)}]$)

Let hyperparameters $(n^{(0)}, y^{(0)})$ vary in a set → set of priors

Bayesian Inference with Sets of Priors

Standard Bayesian inference procedure

prior + likelihood = posterior

using Bayes' Rule

All inferences are based on the posterior

(e.g., point estimate = $\mathrm{E}[\psi \mid n^{(n)}, y^{(n)}]$)

Let hyperparameters $(n^{(0)}, y^{(0)})$ vary in a set → set of priors

Generalised Bayesian inference procedure

set of priors + likelihood = set of posteriors

All inferences are based on the set of posteriors

Coherence (consistency of inferences) ensured by using

Generalised Bayes' Rule (GBR, Walley 1991)

= element-wise application of Bayes' Rule

Set of Priors can be Convex

Convex Set of Priors

$$\mathcal{M}^{(0)} = \text{conv}\Big(\big\{p(\psi \mid n^{(0)}, \mathbf{y}^{(0)}) : (n^{(0)}, \mathbf{y}^{(0)}) \in \qquad \big\}\Big)$$

$\mathcal{M}^{(0)}$ = finite convex mixtures of canonical conjugate priors defined by

Set of Priors can be Convex

Convex Set of Priors

$$\mathcal{M}^{(0)} = \operatorname{conv}\Big(\Big\{p(\psi \mid n^{(0)}, y^{(0)}) : (n^{(0)}, y^{(0)}) \in \qquad \Big\}\Big)$$

$\mathcal{M}^{(0)}$ = finite convex mixtures of canonical conjugate priors defined by

Updating & mixture commute → set of posteriors can be written as. . .

Convex Set of Posteriors

$$\mathcal{M}^{(n)} = \operatorname{conv}\Big(\Big\{p(\psi \mid n^{(n)}, y^{(n)}) : (n^{(n)}, y^{(n)}) \in \qquad \Big\}\Big)$$

where $\quad = \Big\{(n^{(n)}, y^{(n)}) : (n^{(0)}, y^{(0)}) \in \qquad \Big\}.$

$\mathcal{M}^{(n)}$ = finite convex mixtures of canonical conjugate posteriors defined by set of updated hyperparameters

Set of Priors can be Convex

Convex Set of Priors

$$\mathcal{M}^{(0)} = \text{conv}\Big(\big\{p(\psi \mid n^{(0)}, \mathbf{y}^{(0)}) : (n^{(0)}, \mathbf{y}^{(0)}) \in \qquad \big\}\Big)$$

$\mathcal{M}^{(0)}$ = finite convex mixtures of canonical conjugate priors defined by

Updating & mixture commute ➡ set of posteriors can be written as. . .

Convex Set of Posteriors

$$\mathcal{M}^{(n)} = \text{conv}\Big(\big\{p(\psi \mid n^{(n)}, \mathbf{y}^{(n)}) : (n^{(n)}, \mathbf{y}^{(n)}) \in \qquad \big\}\Big)$$

where $= \big\{(n^{(n)}, \mathbf{y}^{(n)}) : (n^{(0)}, \mathbf{y}^{(0)}) \in \qquad \big\}.$

$\mathcal{M}^{(n)}$ = finite convex mixtures of canonical conjugate posteriors defined by set of updated hyperparameters

Convex sets make the procedure very general (mixture distributions), but useful only for inferences that are *linear* in the posteriors (expectations: yes, variances: no)

Generalised Bayesian Inference Procedure

single prior $(n^{(0)}, y^{(0)})$ ➡ set of priors $\mathcal{M}^{(0)}$ (defined via)

$\mathrm{E}[\psi \mid n^{(0)}, y^{(0)}, \boldsymbol{x}]$ ➡ $\big[\underline{\mathrm{E}}[\psi \mid \quad , \boldsymbol{x}], \overline{\mathrm{E}}[\psi \mid \quad , \boldsymbol{x}]\big]$

$P(\psi \in A \mid n^{(0)}, y^{(0)}, \boldsymbol{x})$ ➡ $\big[\underline{\mathrm{P}}[\psi \in A \mid \quad , \boldsymbol{x}], \overline{\mathrm{P}}[\psi \in A \mid \quad , \boldsymbol{x}]\big]$

Lower/upper posterior expectation by min/max over set of posteriors

Generalised Bayesian Inference Procedure

single prior $(n^{(0)}, y^{(0)})$ ➡ set of priors $\mathcal{M}^{(0)}$ (defined via)

$\mathrm{E}[\psi \mid n^{(0)}, y^{(0)}, \boldsymbol{x}]$ ➡ $\left[\underline{\mathrm{E}}[\psi \mid \quad , \boldsymbol{x}], \overline{\mathrm{E}}[\psi \mid \quad , \boldsymbol{x}]\right]$

$P(\psi \in A \mid n^{(0)}, y^{(0)}, \boldsymbol{x})$ ➡ $\left[\underline{\mathrm{P}}[\psi \in A \mid \quad , \boldsymbol{x}], \overline{\mathrm{P}}[\psi \in A \mid \quad , \boldsymbol{x}]\right]$

Lower/upper posterior expectation by min/max over set of posteriors

Interpretation

Shorter intervals ⬌ more precise probability statements

Lottery A
Number of winning tickets:
exactly known as 5 out of 100
➡ $P(\text{win}) = 5/100$

Lottery B
Number of winning tickets:
not exactly known, supposedly
between 1 and 7 out of 100
➡ $P(\text{win}) = [1/100, 7/100]$

Generalised Bayesian Inference Procedure

Shorter intervals ⟷ more precise probability statements

- larger $n^{(0)}$ values as compared to n ⟶ larger
  ⟶ more vague inferences
  (more weight on imprecise prior $\mathcal{M}^{(0)}$ leads to more imprecise posterior $\mathcal{M}^{(n)}$)

Generalised Bayesian Inference Procedure

Shorter intervals ⟷ more precise probability statements

- larger $n^{(0)}$ values as compared to n ➡ larger ➡ more vague inferences
  (more weight on imprecise prior $\mathcal{M}^{(0)}$ leads to more imprecise posterior $\mathcal{M}^{(n)}$)
- larger n as compared to (range of) $n^{(0)}$ ➡ smaller ➡ more precise inferences

Generalised Bayesian Inference Procedure

Shorter intervals $\Longleftrightarrow$ more precise probability statements

- larger $n^{(0)}$ values as compared to n $\Longrightarrow$ larger
  $\Longrightarrow$ more vague inferences
  (more weight on imprecise prior $\mathcal{M}^{(0)}$ leads to more imprecise posterior $\mathcal{M}^{(n)}$)
- larger n as compared to (range of) $n^{(0)}$ $\Longrightarrow$ smaller
  $\Longrightarrow$ more precise inferences
- $n \to \infty$ $\Longrightarrow$ $y^{(n)}$ values in $\quad \to \frac{\tau(\mathbf{x})}{n}$ $\Longrightarrow$ 'Bayesian consistency'

Generalised Bayesian Inference Procedure

Shorter intervals ⟷ more precise probability statements

- larger $n^{(0)}$ values as compared to n ➡ larger
  ➡ more vague inferences
  (more weight on imprecise prior $\mathcal{M}^{(0)}$ leads to more imprecise posterior $\mathcal{M}^{(n)}$)
- larger n as compared to (range of) $n^{(0)}$ ➡ smaller
  ➡ more precise inferences
- $n \to \infty$ ➡ $y^{(n)}$ values in $\to \frac{\tau(\mathbf{x})}{n}$ ➡ 'Bayesian consistency'
- larger range of $y^{(0)}$ in ➡ larger range of $y^{(n)}$ in
  ➡ more vague inferences
  (more imprecise prior $\mathcal{M}^{(0)}$ leads to more imprecise $\mathcal{M}^{(n)}$)

Generalised Bayesian Inference Procedure

- Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$

Generalised Bayesian Inference Procedure

- Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$

Generalised Bayesian Inference Procedure

- Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$
- $\rightarrow$ is easy: $n^{(n)} = n^{(0)} + n$, $y^{(n)} = \frac{n^{(0)}}{n^{(0)}+n} y^{(0)} + \frac{n}{n^{(0)}+n} \frac{\tau(\mathbf{x})}{n}$

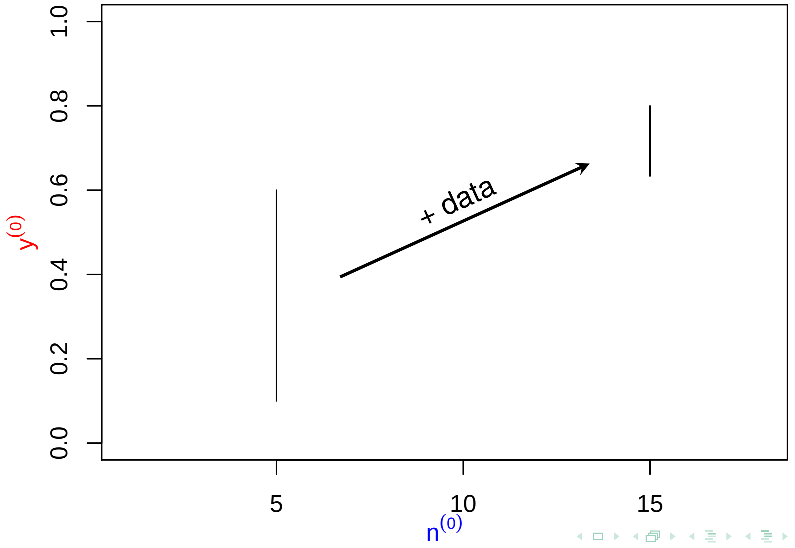
Generalised Bayesian Inference Procedure

- Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$
- $\rightarrow$ is easy: $n^{(n)} = n^{(0)} + n$, $y^{(n)} = \frac{n^{(0)}}{n^{(0)}+n} y^{(0)} + \frac{n}{n^{(0)}+n} \frac{\tau(\mathbf{x})}{n}$
- Quantities linear in $p(\psi \mid n^{(n)}, \mathbf{y}^{(n)})$ (e.g., $\mathrm{E}[g(\psi) \mid n^{(n)}, \mathbf{y}^{(n)}]$):
  ➡ bounds attained at “pure” posteriors $p(\psi \mid n^{(n)}, \mathbf{y}^{(n)})$
  ➡ straighforward to calculate: optimise over only

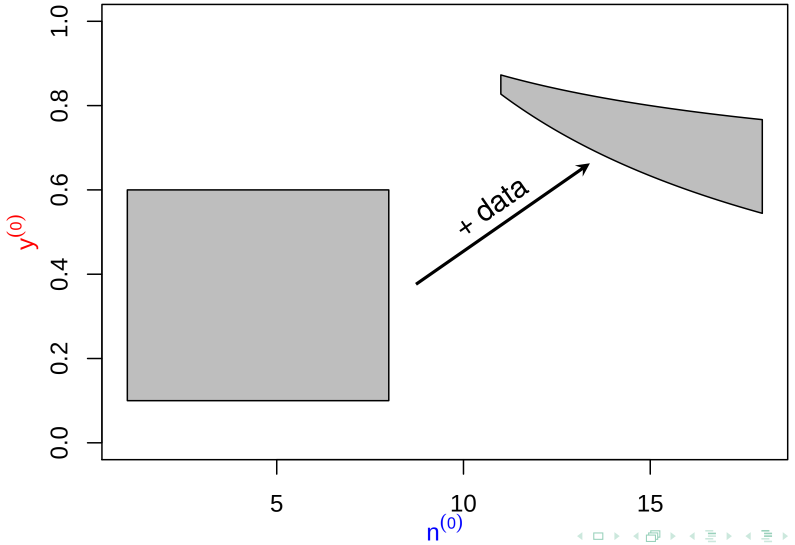
Generalised Bayesian Inference Procedure

- Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$
- $\rightarrow$ is easy: $n^{(n)} = n^{(0)} + n$, $y^{(n)} = \frac{n^{(0)}}{n^{(0)}+n} y^{(0)} + \frac{n}{n^{(0)}+n} \frac{\tau(\mathbf{x})}{n}$
- Quantities linear in $p(\psi \mid n^{(n)}, \mathbf{y}^{(n)})$ (e.g., $\operatorname{E}[g(\psi) \mid n^{(n)}, \mathbf{y}^{(n)}]$):
  ➡ bounds attained at “pure” posteriors $p(\psi \mid n^{(n)}, \mathbf{y}^{(n)})$
  ➡ straightforward to calculate: optimise over only
- Often, optimising over $(n^{(n)}, \mathbf{y}^{(n)}) \in$ is also easy: posterior ‘guess’ for $\frac{\tau(\mathbf{x})}{n}$ (think: $\bar{\mathbf{x}}$) = $y^{(n)}$
  ➡ closed form solution given has ‘nice’ shape

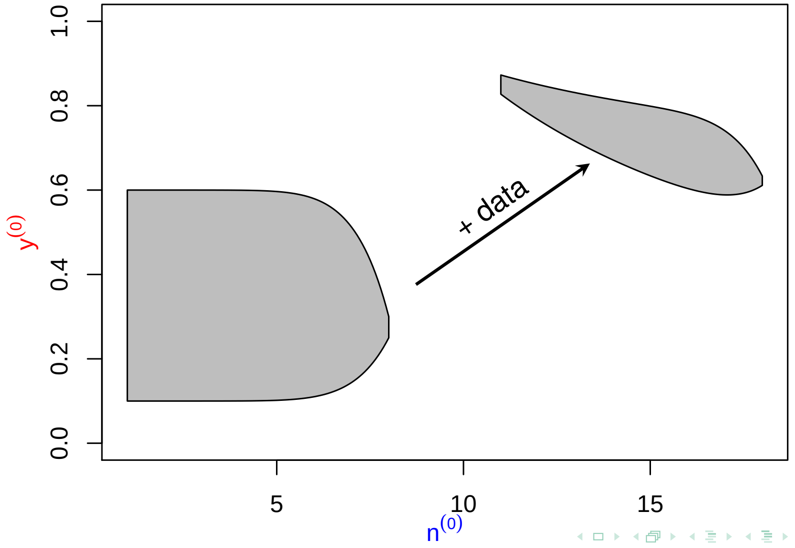
Parameter Set Shapes



Parameter Set Shapes



Parameter Set Shapes



Parameter Set Shapes

- Shape of influences shape of

Parameter Set Shapes

- Shape of influences shape of
- Shape of influences model behaviour
  → shape of is a crucial modelling choice

Parameter Set Shapes

- Shape of influences shape of
- Shape of influences model behaviour
  ➡ shape of is a crucial modelling choice
- $= [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$ (*rectangle*) is very easy to elicit and gives good model behaviour for prior-data conflict

Parameter Set Shapes

- Shape of influences shape of
- Shape of influences model behaviour
  → shape of is a crucial modelling choice
- $= [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$ (*rectangle*) is very easy to elicit and gives good model behaviour for prior-data conflict

Prior-Data Conflict

- *informative prior beliefs* and *trusted data* (sampling model correct, no outliers, etc.) are in conflict
- "[...] the prior [places] its mass primarily on distributions in the sampling model for which the observed data is surprising" (Evans & Moshonov, 2006)
- there are not enough data to overrule the prior

Prior-Data Conflict: Example

- Bernoulli observations: 0/1 observations (team wins no/yes)

Prior-Data Conflict: Example

- Bernoulli observations: 0/1 observations (team wins no/yes)
- given: a set of observations (team won 12 out of 16 matches)

Prior-Data Conflict: Example

- Bernoulli observations: 0/1 observations (team wins no/yes)
- given: a set of observations (team won 12 out of 16 matches)
- additional to observations, we have strong prior information (we are convinced that $P(\text{win})$ should be around 0.75)

Prior-Data Conflict: Example

- Bernoulli observations: 0/1 observations (team wins no/yes)
- given: a set of observations (team won 12 out of 16 matches)
- additional to observations, we have strong prior information (we are convinced that $P(\text{win})$ should be around 0.75)
- we are, e.g., interested in (predictive) probability P that team wins in the next match

Prior-Data Conflict: Example

- Bernoulli observations: 0/1 observations (team wins no/yes)
- given: a set of observations (team won 12 out of 16 matches)
- additional to observations, we have strong prior information (we are convinced that P(win) should be around 0.75)
- we are, e.g., interested in (predictive) probability P that team wins in the next match

Beta-Binomial Model

Data :	$s \mid p$	$\sim$	$\text{Binom}(p)$
conjugate prior:	$p \mid n^{(0)}, y^{(0)}$	$\sim$	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid n^{(n)}, y^{(n)}$	$\sim$	$\text{Beta}(n^{(n)}, y^{(n)}) \quad (\frac{\tau(\boldsymbol{x})}{n} = \frac{s}{n})$

where s = number of wins in the n matches observed

Beta-Binomial Model

Beta-Binomial Model

Data :	$s \mid p$	$\sim$	$\text{Binom}(p)$
conjugate prior:	$p \mid n^{(0)}, y^{(0)}$	$\sim$	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid n^{(n)}, y^{(n)}$	$\sim$	$\text{Beta}(n^{(n)}, y^{(n)}) \quad (\frac{\tau(\boldsymbol{x})}{n} = \frac{s}{n})$

Beta-Binomial Model

Beta-Binomial Model

Data :	$s \mid p$	$\sim$	$\text{Binom}(p)$
conjugate prior:	$p \mid n^{(0)}, y^{(0)}$	$\sim$	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid n^{(n)}, y^{(n)}$	$\sim$	$\text{Beta}(n^{(n)}, y^{(n)}) \quad (\frac{\tau(\boldsymbol{x})}{n} = \frac{s}{n})$

$$P = \text{E}[p \mid n^{(n)}, y^{(n)}]$$

Beta-Binomial Model

Beta-Binomial Model

Data :	$s \mid p$	$\sim$	$\text{Binom}(p)$
conjugate prior:	$p \mid n^{(0)}, y^{(0)}$	$\sim$	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid n^{(n)}, y^{(n)}$	$\sim$	$\text{Beta}(n^{(n)}, y^{(n)})$ $\quad (\frac{\tau(\boldsymbol{x})}{n} = \frac{s}{n})$

$$P = \text{E}[p \mid n^{(n)}, y^{(n)}] = y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

Beta-Binomial Model

Beta-Binomial Model

Data :	$s \mid p$	$\sim$	$\text{Binom}(p)$
conjugate prior:	$p \mid n^{(0)}, y^{(0)}$	$\sim$	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid n^{(n)}, y^{(n)}$	$\sim$	$\text{Beta}(n^{(n)}, y^{(n)})$ $\quad (\frac{\tau(\mathbf{x})}{n} = \frac{s}{n})$

$$P = \text{E}[p \mid n^{(n)}, y^{(n)}] = y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$$n^{(n)} = n^{(0)} + n$$

Beta-Binomial Model

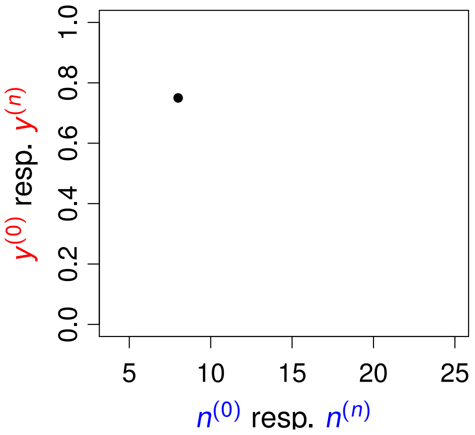
Beta-Binomial Model

$$\begin{array}{r|lcl} \text{Data:} & s \mid p & \sim & \text{Binom}(p) \\ \text{conjugate prior:} & p \mid n^{(0)}, y^{(0)} & \sim & \text{Beta}(n^{(0)}, y^{(0)}) \\ \hline \text{posterior:} & p \mid n^{(n)}, y^{(n)} & \sim & \text{Beta}(n^{(n)}, y^{(n)}) \quad (\frac{\tau(\boldsymbol{x})}{n} = \frac{s}{n}) \end{array}$$

$$P = \text{E}[p \mid n^{(n)}, y^{(n)}] = y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$$n^{(n)} = n^{(0)} + n \qquad \text{Var}(p \mid n^{(n)}, y^{(n)}) = \frac{y^{(n)}(1 - y^{(n)})}{n^{(n)} + 1}$$

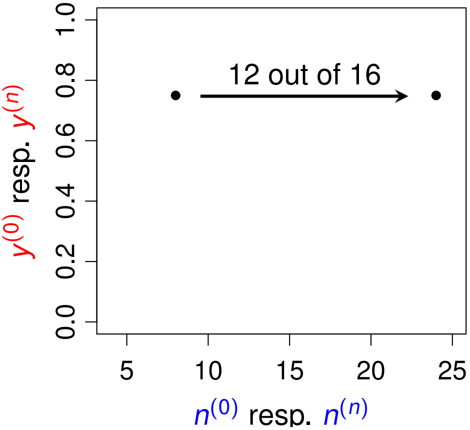
Beta-Binomial Model



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

Beta-Binomial Model



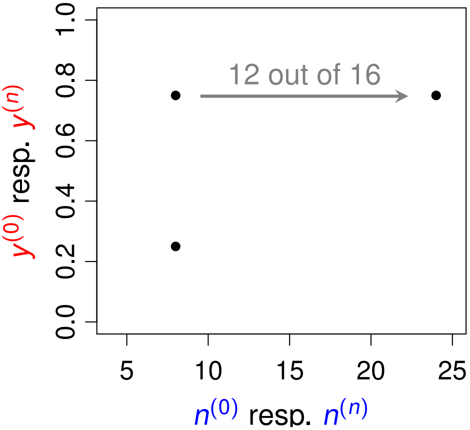
no conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

▼

$n^{(n)} = 24$, $y^{(n)} = 0.75$

Beta-Binomial Model



no conflict:

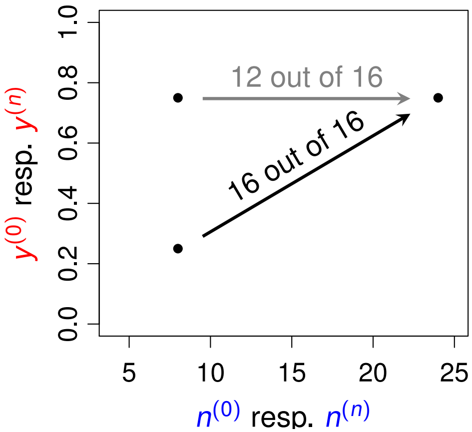
prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

$n^{(n)} = 24$, $y^{(n)} = 0.75$

prior-data conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.25$
data $s/n = 16/16 = 1$

Beta-Binomial Model



no conflict:

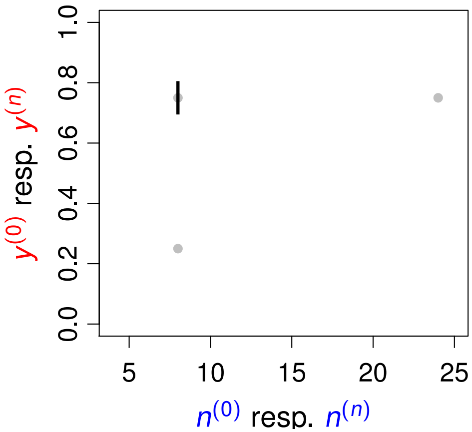
prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

$n^{(n)} = 24$, $y^{(n)} = 0.75$

prior-data conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.25$
data $s/n = 16/16 = 1$

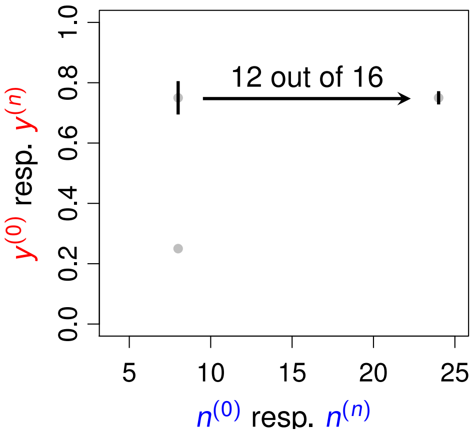
Imprecise BBM with $n^{(0)}$ fixed = IDM (Walley 1996)



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

Imprecise BBM with $n^{(0)}$ fixed = IDM (Walley 1996)



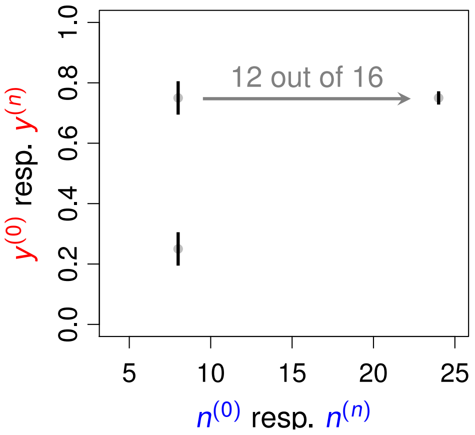
no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

▼

$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

Imprecise BBM with $n^{(0)}$ fixed = IDM (Walley 1996)



no conflict:

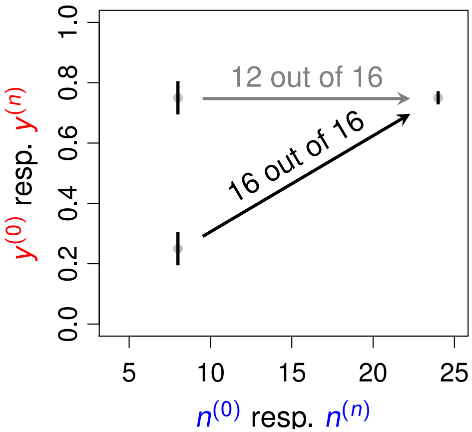
prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

prior-data conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.2, 0.3]$
data $s/n = 16/16 = 1$

Imprecise BBM with $n^{(0)}$ fixed = IDM (Walley 1996)



no conflict:

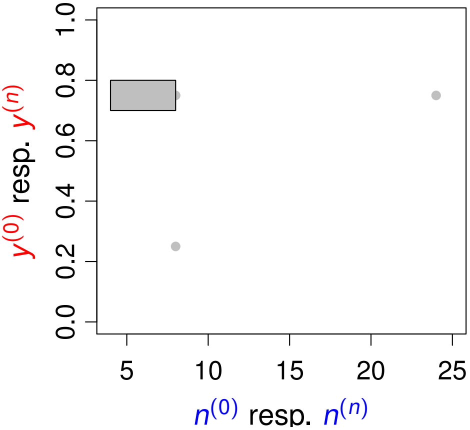
prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

prior-data conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.2, 0.3]$
data $s/n = 16/16 = 1$

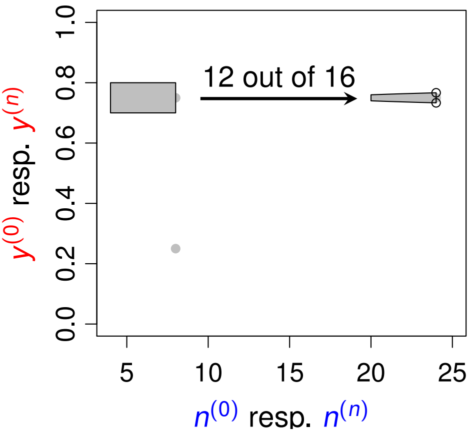
Imprecise BBM with $[\underline{n}^{(0)}, \overline{n}^{(0)}]$ (Walley 1991, §5.4.3)



no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

Imprecise BBM with $[\underline{n}^{(0)}, \overline{n}^{(0)}]$ (Walley 1991, §5.4.3)

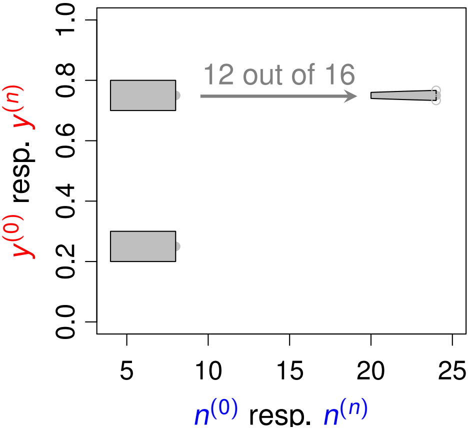


no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

$y^{(n)} \in [0.73, 0.77]$

Imprecise BBM with $[\underline{n}^{(0)}, \overline{n}^{(0)}]$ (Walley 1991, §5.4.3)



no conflict:

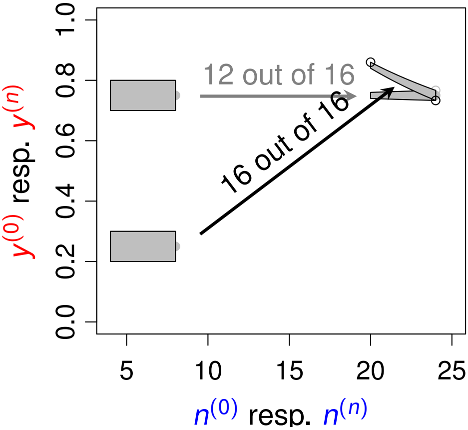
prior $n^{(0)} \in [4,8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

$y^{(n)} \in [0.73, 0.77]$

prior-data conflict:

prior $n^{(0)} \in [4,8]$, $y^{(0)} \in [0.2, 0.3]$
data $s/n = 16/16 = 1$

Imprecise BBM with $[\underline{n}^{(0)}, \overline{n}^{(0)}]$ (Walley 1991, §5.4.3)



no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

▼

$y^{(n)} \in [0.73, 0.77]$

prior-data conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.2, 0.3]$
data $s/n = 16/16 = 1$

▼

$y^{(n)} \in [0.73, 0.86]$

Parameter Set Shapes: Discussion

- $= n^{(0)} \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$:
  IDM (Walley 1996), Quaghebeur & de Cooman (2005)
  - posterior parameter set has same form $= n^{(n)} \times [\underline{y}^{(n)}, \overline{y}^{(n)}]$
  - optimise over $[\underline{y}^{(n)}, \overline{y}^{(n)}]$ only
  - no prior-data conflict reaction: same imprecision as without conflict (just like precise priors)

Parameter Set Shapes: Discussion

- $= n^{(0)} \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$:
  IDM (Walley 1996), Quaghebeur & de Cooman (2005)
  - posterior parameter set has same form $= n^{(n)} \times [\underline{y}^{(n)}, \overline{y}^{(n)}]$
  - optimise over $[\underline{y}^{(n)}, \overline{y}^{(n)}]$ only
  - no prior-data conflict reaction: same imprecision as without conflict (just like precise priors)
- $= [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$: Walley (1991, §5.4.3), *generalized iLUCK-models* (Walter & Augustin 2009)
  - still simple to elicit, allows flexible weighing of prior and data
  - additional imprecision in case of prior-data conflict
    ⟶ more cautious inferences
  - have non-trivial forms (banana / spotlight)
  - however, closed form for min / max $y^{(n)}$ over
  - general optimisation over more difficult, but doable
  - **R** package `luck`: do optimisation over actually

Parameter Set Shapes: Discussion

- Need a range of $n^{(0)}$ values for prior-data conflict reaction

Parameter Set Shapes: Discussion

- Need a range of $n^{(0)}$ values for prior-data conflict reaction
- Other set shapes are possible, but may be more difficult to elicit

Parameter Set Shapes: Discussion

- Need a range of $n^{(0)}$ values for prior-data conflict reaction
- Other set shapes are possible, but may be more difficult to elicit
- Prior information may be such that range of $y^{(0)}$ changes with $n^{(0)}$ (or vice versa)

Parameter Set Shapes: Discussion

- Need a range of $n^{(0)}$ values for prior-data conflict reaction
- Other set shapes are possible, but may be more difficult to elicit
- Prior information may be such that range of $y^{(0)}$ changes with $n^{(0)}$ (or vice versa)
- *Near-ignorance priors*: such that prior inferences are *vacuous*, but posterior inferences are informative
  - IDM (Walley 1996): range of $y_j^{(0)} = (0,1)\ \forall j$
  - Benavoli & Zaffalon (2012): range of $y^{(0)} = (-\infty, +\infty)$ while $\overline{n}^{(0)}$ decreasing with $y^{(0)}$ (to avoid $n^{(0)}|y^{(0)}| = \infty$, i.e. vacuous posterior inferences)

Parameter Set Shapes: Outlook

Work in progress (joint work with Frank Coolen):
parameter set shape enabling. . .

- additional imprecision in case of prior-data conflict (as before)
- less imprecision for *strong prior-data agreement*

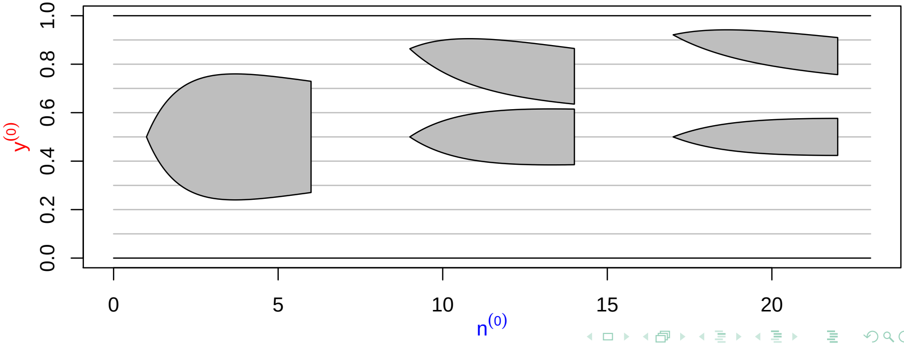
via a different parametrisation of priors suggested by Mik Bickis

Parameter Set Shapes: Outlook

Work in progress (joint work with Frank Coolen):
parameter set shape enabling. . .

- additional imprecision in case of prior-data conflict (as before)
- less imprecision for *strong prior-data agreement*

via a different parametrisation of priors suggested by Mik Bickis



Conclusion

- Conjugate priors are a convenient tool for Bayesian inference but there are some pitfalls
  - Hyperparameters are easy to interpret and elicit
  - Averaging property makes calculations simple,
    but inadequate model behaviour in case of prior-data conflict

Conclusion

- Conjugate priors are a convenient tool for Bayesian inference but there are some pitfalls
  - Hyperparameters are easy to interpret and elicit
  - Averaging property makes calculations simple, but inadequate model behaviour in case of prior-data conflict
- Sets of conjugate priors maintain advantages & mitigate issues
  - Hyperparameter set shape is important
  - Reasonable choice: *rectangular* $= [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$
  - Bounds for hyperparameters easy to interpret and elicit
  - Additional imprecison in case of prior-data conflict leads to cautious inferences if, and only if, caution is needed
  - Shape for less imprecision in case of strong prior-data agreement is in development

References

Atwood, C. L. (1996). "Constrained noninformative priors in risk assessment". In: *Reliability Engineering and System Safety* 53, pp. 37–46.

Evans, M. and H. Moshonov (2006). "Checking for Prior-Data Conflict". In: *Bayesian Analysis* 1, pp. 893–914.

Good, I. J. (1965). *The estimation of probabilities*. Cambridge (MA): MIT Press.

Høyland, Arnljot and Marvin Rausand (1994). *System reliability theory: models and statistical methods*. A Wiley interscience publication. New York, NY: Wiley. ISBN: 0-471-59397-4.

Kelly, Dana and Corwin Atwood (2011). "Finding a minimally informative Dirichlet prior distribution using least squares". In: *Reliability Engineering and System Safety* 96.3, pp. 398–402. ISSN: 0951-8320. DOI: 10.1016/j.ress.2010.11.008.

Quaeghebeur, E. and G. de Cooman (2005). "Imprecise probability models for inference in exponential families". In: *ISIPTA '05*. Ed. by F. Cozman, R. Nau, and T. Seidenfeld. Manno: SIPTA, pp. 287–296.

Troffaes, Matthias C. M., Gero Walter, and Dana L. Kelly (2013). *A Robust Bayesian Approach to Modelling Epistemic Uncertainty in Common-Cause Failure Models*. Available at http://arxiv.org/abs/1301.0533. Submitted to: Reliability Engineering & System Safety.

Walley, Peter (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.

Walley, Peter (1996). "Inferences from multinomial data: Learning about a bag of marbles". In: *Journal of the Royal Statistical Society, Series B* 58.1, pp. 3–34.

Walter, Gero and Thomas Augustin (2009). "Imprecision and Prior-data Conflict in Generalized Bayesian Inference". In: *Journal of Statistical Theory and Practice* 3, pp. 255–271.